The law of everything.

Purtova, Nadezhda

Published in:
Law, Innovation and Technology

DOI:
10.1080/17579961.2018.1452176

Publication date:
2018

Routledge
Taylor & Francis Group

Law, Innovation and Technology

ISSN: 1757-9961 (Print) 1757-997X (Online) Journal homepage: http://www.tandfonline.com/loi/rlit20

The law of everything. Broad concept of personal data and future of EU data protection law

Nadezhda Purtova

To cite this article: Nadezhda Purtova (2018) The law of everything. Broad concept of personal data and future of EU data protection law, Law, Innovation and Technology, 10:1, 40-81, DOI: 10.1080/17579961.2018.1452176

To link to this article: https://doi.org/10.1080/17579961.2018.1452176

LAW, INNOVATION AND TECHNOLOGY, 2018
VOL. 10, NO. 1, 40–81
https://doi.org/10.1080/17579961.2018.1452176

Routledge
Taylor & Francis Group

The law of everything. Broad concept of personal data and future of EU data protection law

Nadezhda Purtova

Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, Tilburg, The Netherlands

ABSTRACT

Article 29 Working Party guidelines and the case law of the CJEU facilitate a plausible argument that in the near future everything will be or will contain personal data, leading to the application of data protection to everything: technology is rapidly moving towards perfect identifiability of information; datafication and advances in data analytics make everything (contain) information; and in increasingly 'smart' environments any information is likely to relate to a person in purpose or effect. At present, the broad notion of personal data is not problematic and even welcome. This will change in future. When the hyperconnected onlife world of data-driven agency arrives, the intensive compliance regime of the General Data Protection Regulation (GDPR) will become 'the law of everything', well-meant but impossible to maintain. By then we should abandon the distinction between personal and non-personal data, embrace the principle that all data processing should trigger protection, and understand how this protection can be scalable.

ARTICLE HISTORY Received 8 September 2017; Accepted 19 February 2018

KEYWORDS Personal data; GDPR; material scope; information relating to; *Breyer*; *Nowak*

1. Introduction

In *The Morality of Law*, Lon Fuller tells a tale of a young ruler who undertakes to reform the law of the land. After a few attempts at it, all met by public discontent, the ruler wants to teach his subjects a lesson and makes it a crime 'to cough, sneeze, hiccough, faint or fall down in the presence of the king … [and] not to understand, believe in, and correctly profess the doctrine of evolutionary, democratic redemption'.[1] Unsurprisingly, the citizens threaten to disregard the new law, since '[t]o command what cannot be done is not to make law; it is to

CONTACT Nadezhda Purtova N.N.Purtova@uvt.nl Tilburg Institute for Law, Technology, and Society (TILT), Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The Netherlands

[1]LL Fuller, *The Morality of Law* (Yale University Press, 1969), 36–37.

unmake law, for a command that cannot be obeyed serves no end but confusion, fear and chaos'.[2] This can as well become the future of European data protection.

This paper will argue that the material scope of EU data protection law, specifically, the General Data Protection Regulation[3] ('GDPR'), is growing so broad that the good intentions to provide the most complete protection possible are likely to backfire in a very near future, resulting in system overload. The concept 'personal data' determining the material scope of data protection is meant to be broad but is bound to expand even further and as a result to apply to an exponentially growing range of situations. This is due to the in-built possibilities for the evolving interpretation of the concept itself, exploding generation and aggregation of data, as well as advances in data analytics. As our environment is rapidly approaching what some call 'onlife'[4] where our daily existence is mediated by information technology, everything in this environment – weather, waste water, exam scripts – is being increasingly 'datified', and literally any data can be plausibly argued to be personal. Four major transformations are at the core of this shift:

a. the blurring of the distinction between reality and virtuality;
b. the blurring of the distinction between human, machine and nature;
c. the reversal from information scarcity to information abundance; and
d. the shift from the primacy of stand-alone things, properties and binary relations, to the primacy of interactions, processes and networks.[5]

As a result, European data protection law is facing a risk of becoming 'the law of everything', meant to deliver the highest legal protection under all circumstances, but in practice impossible to comply with and hence ignored or discredited as conducive to abuse of rights and unreasonable.

Discussion of the expanding scope of personal data has been brewing for quite some time in the data protection community. Many privacy and data protection scholars have been critical of the concept of personal data as growing too broad. The mainstream literature is centred on one element of the concept of personal data, ie identifiability of a person, and the corresponding strand of technological development, ie re-identification and de-anonymisation algorithms. To name just a few key authors arguing to that effect, the works of Ohm,[6] Sweeney[7] and Schwartz and

[2]Ibid.

[3]Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), published in *Official Journal of the European Union*, L 119, 4 May 2016.

[4]The term 'onlife' was coined by Luciano Floridi and refers to 'the new experience of a hyperconnected reality within which it is no longer sensible to ask whether one may be online or offline' (Luciano Floridi, 'Introduction' in Luciano Floridi (ed), *The Online Manifesto. Being Human in a Hyperconnected Era* (Springer, 2015), 1).

[5]Ibid, 2.

[6]P Ohm, 'Broken Promises of Privacy' (2010) 57 *UCLA L. Rev*. 1701, 1742 et seq. and 1759.

[7]L Sweeney, 'Simple Demographics Often Identify People Uniquely' (2000) Carnegie Mellon University, Data Privacy Working Paper 3 http://dataprivacylab.org/projects/identifiability/paper1.pdf (accessed 18 February 2018).

Solove[8] suggest that, given the progress of data processing technologies and the amount of data available for analysis, absolute and irreversible anonymity is no longer possible. Tene and Polonetsky note how Big Data analytics makes a binary distinction between identifiable and non-identifiable information meaningless.[9] Schwartz and Solove propose to keep personal data (or personally identifiable information, a functional equivalent of personal data in the USA) as a threshold of protection, but with a sharper definition, namely, one based on the risk of identification from '0' (zero risk of identification) to 'identified', and to treat information with varying degrees of identifiability differently.[10] Alongside these suggestions, we should note that the EU Court of Justice has recently adopted a broad approach to identifiability in the *Breyer* case.[11]

What these authors seem to overlook, however, is that the problem with the concept of personal data goes beyond simple identifiability, because the second essential element of the concept of 'personal data', ie the relation of information to a person, is problematic as well.

As I will argue in this paper, in the age of the Internet of Things, datafication, advanced data analytics and data-driven decision-making, any information relates to a person in the sense of European data protection law. From the perspective of the present and near future of data protection, I welcome this broad interpretation of personal data and its adoption in the data protection practice. To be clear, my argument is not made in support of a narrower concept of personal data, eg akin to personally identifiable information in the US law.[12] Nor do I defend a narrower scope of data protection. Indeed, if all data has a potential to impact people and is therefore personal, all data should trigger some sort of protection against possible negative impacts.[13] The broad interpretation of the concept 'personal data' and the resulting broad legal protection are not the core of the problem as I see it. The problem is that in the circumstances where all data is personal and triggers data protection, a highly intensive and non-scalable regime of rights and obligations that results from the GDPR cannot be upheld in a meaningful way. From the perspective of a more long-term future of data protection law, while I still support the broad scope of the legal protection that a broad notion of personal data brings, I want to enter a caution about the

[8]P Schwartz and D Solove, 'The PII Problem: Privacy and a New Concept of Personally Identifiable Information' (2011) 86 *N.Y.U. L. Rev.* 1814, 1877.
[9]O Tene and J Polonetsky, 'Big Data for All: Privacy and User Control in the Age of Analytics' (2013) 11 *Northwestern Journal of Technology and Intellectual Property* 258.
[10]Schwartz and Solove (n 8).
[11]Case C-582/14, *Patrick Breyer v. Bundesrepublik Deutschland* [2016] ECLI:EU:C:2016:779.
[12]The scope of the notion of personally identifiable information is discussed and criticised at length in Paul Schwartz and Daniel Solove 'Reconciling Personal Information in the United States and European Union' (2014) 102 *California Law Review* 877.
[13]But see GJ Zwenne, *Diluted Privacy Law. Inaugural Lecture* (University of Leiden, 2013) 33 et seq.

consequences of imposing the same high intensity of obligations in all data processing situations.

I will start the argument by breaking up the definitions of personal data under the 1995 Data Protection Directive[14] ('DPD') and the GDPR to demonstrate how they are inherently flexible and capable of being stretched to accommodate changing contexts (Part 2). I will then review how the concept of personal data has been applied so far, mainly referring to the Article 29 Working Party ('WP29') opinion on the concept of personal data[15] ('WP 136'). The WP136 opinion will serve as a mental scaffolding around which the paper's argument will be constructed. The reason is that albeit not formally binding, the WP136 possesses undeniable 'persuasive authority' and provides the most comprehensive guidelines for data controllers as to how they should apply the concept of personal data in their day-to-day practice. I will show how the notion of personal data according to the WP136, broadly defined and in light of the rapidly advancing technology, potentially renders everything personal data and subject to the data protection regime (Part 3). Yet, only the CJEU has authority to decide how the definition of 'personal data' should be interpreted under EU data protection law. Therefore, I will review the CJEU's case law and examine whether the Court shares the WP29's broad reading of the concept 'personal data' (Part 4). I will discuss implications of the analysis for the present and future of data protection law in Europe (Part 5), followed by a conclusion (Part 6).

2. Definition of personal data in EU data protection law: flexibility, adaptability, and uncertainty

'Personal data' is one of the key notions of data protection law determining the material scope of the DPD and the GDPR. Only when personal data is processed do the data protection principles, rights and obligations apply (Article 3(1) DPD and Article 2(1) GDPR).

Under the GDPR, which closely follows the DPD,[16] 'personal data' means

> any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.

Anonymous data is the opposite of personal data and refers to information that does not relate to an identified or identifiable person, or to personal

[14]Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, *Official Journal* L 281, 23/11/1995, 0031–0050.
[15]Article 29 Working Party opinion 4/2007 on the concept of personal data, 20 June 2007 ('WP 136').
[16]Article 4(1) GDPR.

data 'rendered anonymous in such a manner that the data subject is not or no longer identifiable'.[17] Processing anonymous data does not trigger application of data protection law. Pseudonymous data, ie personal data after pseudonymisation, is still information pertaining to an identifiable person and subject to data protection law.[18]

The resulting definition of personal data is broad, flexible, and adaptable to technological context.[19] The references to '*identifiable* natural person' and 'information *relating to* a natural person' invite interpretation as to what constitutes a relevant possibility of identification and a relevant relationship between information and an individual.

Recital 26 GDPR adopts a test of *reasonable likelihood* of identification 'by the controller or by another person', taking into account not the subjective ability to identify, but the state of art of technology at the time of processing:

> To ascertain whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments.

As Schwartz and Solove point out, Recital 26 GDPR makes the GDPR concept of 'personal data' suitable for 'a tailored, context-specific analysis for deciding whether or not personal data is present'.[20] The same piece of data can be anonymous at the time of collection, but turn into personal later, just sitting there, simply by virtue of technological progress.

In addition to identifiability, 'relate to' is another element of the definition that invites context-dependent assessment. To establish the status of information as personal data, whether or not this information relates to a person has to be considered first, even prior to the identifiability analysis. Like 'identifiability', 'information relating to' a natural person may be interpreted broadly and narrowly, and invites a judgement on what kind and degree of relationship of information to a person is significant, as well as whether this relationship is present under particular circumstances. Neither the DPD nor the GDPR gives any guidelines as to how 'relating to' is to be understood.

Due to the flexibility of the definition in the DPD, and despite the DPD being an instrument of complete harmonisation, there has been a significant

[17] Recital 26 DPD and Recital 26 GDPR.
[18] Recital 26 GDPR; 'pseudonymisation' means 'the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person' (Article 4(5) GDPR).
[19] Schwartz and Solove (n 12), 9 (ssrn pre-print page numbering).
[20] Schwartz and Solove (n 12), 9 (ssrn page numbering).

divergence in how the national legislation has so far implemented the definition of personal data;[21] the complete harmonisation only precludes Member States from adding additional elements to the harmonised provisions, but still allows 'deciding the details or choosing between options'.[22] Also, the Recitals detailing and clarifying the definition of personal data are non-binding, and the Member States are under no obligation to include those in their national implementation. It remains to be seen if the choice of a regulation as a legislative instrument will result in a uniform application of the GDPR definition of personal data across the EU member states.

3. Weather as personal data? Article 29 working party opinion

3.1. Introducing the WP 136: streamlining national implementation and relevance after May 2018

To streamline the national implementation of the definition of personal data, the Article 29 Working Party, an EU advisory authority on the matters of data protection, adopted a non-binding opinion on the concept of personal data.[23] Even though the WP29 opinion concerns the concept of personal data in the DPD, it will most likely remain significant for data protection compliance after the GDPR becomes effective, since, as Advocate General Kokott has observed, 'the latter will not affect the concept of personal data'.[24]

The concept 'personal data' – as interpreted in the WP136 – lives up to its potential to become an all-encompassing notion. The WP29 begins with policy considerations, the most relevant of which for this analysis are that the notion of personal data is broad, and intends to cover all information which may be linked to an individual; the DPD is sufficiently flexible to strike a balance between the data subjects' rights and legitimate interests of others; while the DPD aims to protect individuals, the scope of data protection rules should not be overstretched at the risk of 'ending up applying data protection rules to situations which were not intended to be covered by those rules and for which they were not designed by the legislator'; at the same time, unduly restrictive interpretation of personal data should be avoided 'so that it can anticipate evolutions and catch all "shadow zones" within its scope'.[25]

WP29 breaks up the definition of personal data into four elements. Personal data is: (a) information, (b) relating to (c) an identified or identifiable

[21]WP 136 (n 15), 3. The UK implementation of the definition is one of the most restrictive: see Christopher Millard and W Kuan Hon, 'Defining 'Personal Data' in e-Social Science' (2012) 15(1) *Information, Communication and Society* 66 http://ssrn.com/abstract=1809182 (accessed 18 February 2018).
[22]Viviane Reding, 'The European data Protection Framework for the Twenty-First Century' (2012) 2(3) *International Data Privacy Law* 121; Case C-468/10 *ASNEF*, ECLI:EU:C:2011:777 [35].
[23]WP 136 (n 15).
[24]Case C-434/16 *Peter Nowak v Data Protection Commissioner* [2017] ECLI:EU:C:2017:994, Opinion of Advocate General Kokott [3].
[25]WP 136 (n 15), 4–5.

(d) natural person. I will only consider the first three as relevant for the argument. Pointing to the flaws of the 'identifiability' element is less controversial compared to the other two, so I will start with it.

3.2. 'Identified or identifiable [natural person]'

The WP29 adopts a broad understanding of what 'identified or identifiable' means. 'Identified' refers to a person who is known, or distinguished in a group, and 'identifiable' is a person who is not identified yet, but identification is possible.[26] One is *directly* identified or identifiable most commonly by reference to a name, in combination with additional information if the name is not unique;[27] one is '*indirectly identifiable*' by the so-called 'unique combinations' of not unique identifiers that allow the individual to be singled out.[28]

The standard for the relevant possibility of identification adopted by the WP29 is whether or not the means of identification are '*reasonably likely to be used*', as under Recital 26 DPD and the similarly phrased Recital 26 GDPR. The WP29 follows the language of the Recital closely, restating that the means of identification are 'reasonably likely to be used *by the controller or any other person*', often interpreted as *by anybody*,[29] which is a significantly broader interpretation allowing for more data to be considered personal, as opposed to the narrow '*by the controller*'. The former approach is often called 'absolute', or 'objective', and the latter 'relative'.[30]

At the same time, the WP29 clarifies that a 'purely hypothetical possibility' of identification is insufficient to meet the standard of 'reasonably likely'.[31] Instead, 'all the factors at stake' should be considered to assess this possibility.[32] Examples of such factors are:

- the cost of identification;
- the intended explicit or implied purpose of processing (when 'the processing … only makes sense if it allows identification of specific individuals and treatment of them in a certain way'[33], the availability of tools of identification should be presumed reasonably likely);

[26]Ibid, 12.
[27]Ibid, 13.
[28]Ibid, 13–14. Later the WP29 recognised cookies as unique identifiers (Article 29 Working Party, 'Opinion 16/2011 on EASA/IAB Best Practice Recommendation on Online Behavioural Advertising' (WP 188) 8 December 2011, 8). For more on 'singling out' see FJ Zuiderveen Borgesius, 'Singling Out People Without Knowing Their Names – Behavioural Targeting, Pseudonymous Data, and the new Data Protection Regulation' (2016) 32 *Computer Law & Security Review* 256–271.
[29]M Taylor, *Genetic Data and the Law: A Critical Perspective on Privacy Protection* (Cambridge University Press, 2012) 140; WG Urgessa, 'The Protective Capacity of the Criterion of Identifiability under EU Data Protection Law' (2016) 2 *Eur. Data Prot. L. Rev.* 521, 529.
[30]Eg Zuiderveen Borgesius (n 28) 264.
[31]WP 136 (n 15), 15.
[32]Ibid, 15–16.
[33]Ibid, 16.

- the risk of organisational dysfunctions (eg breaches of confidentiality duties) and technical failures, including data breaches;
- the state of the art in technology at the time of the processing, including possible technological developments in future, within the lifetime of processing;
- measures to prevent data identification (ie to maintain anonymity) are of importance as a means of avoiding processing personal data altogether, rather than in fulfilment of data security obligations under DPD.[34]

The resulting standard of the reasonable likelihood of identification is quite broad and context-dependent, leading to one major consequence: the status of data as 'personal' is dynamic, ie the same dataset may not obviously be personally identifiable at the start of processing, or from the perspective of the controller, given the tools and data available to him, but become, or appear to have been all along, identifiable from the perspective of another person or once the circumstances change.[35]

It has become a widely accepted point in the privacy and data protection literature that as the data processing technologies advance, and the pool of data which can be combined grows, and as combining databases becomes daily practice of intelligence agencies, 'smart city' municipalities, and advertising, so does the reasonable likelihood of somebody being able to link any piece of information to a person. To name just a few key authors arguing to that effect, I refer to the works of Ohm,[36] Sweeney,[37] Schwartz and Solove,[38] and Tene and Polonetsky.[39]

Data processing practice is rich in examples where data first considered anonymous was identified. In 2000 the combination of a ZIP code, date of birth, and gender was enough to identify 87% of the US population.[40] Famously, film rating records of 500,000 Netflix subscribers were re-identified in 2008 using the openly accessible Internet Movie Database.[41] In 2013 travel routes of celebrities such as Bradley Cooper and Olivia Munn, including street addresses, and whether or not they left a tip, were deduced from the 'anonymised' public database of the New York taxi rides which contained no passenger information, and paparazzi pictures.[42] In 2014 knowing the location of credit card holders on four occasions allowed for the re-identification of 90%

[34]Ibid, 17.
[35]BJ Koops, 'The Trouble with European Data Protection Law' (2014) 4(4) *International Data Privacy Law* 250.
[36]Ohm (n 6), 1742 et seq. and 1759.
[37]Sweeney (n 7).
[38]Schwartz and Solove (n 8), 1877.
[39]Tene and Polonetsky (n 9), 258.
[40]Sweeney (n 7).
[41]A Narayanan and V Shmatikov, 'Robust De-Anonymisation of Large Datasets. (How to Break Anonymity of Netflix Prize Dataset)' (2008) *Proceedings – IEEE Symposium on Security and Privacy* 111–125.
[42]JK Trotter, 'Public NYC Taxicab Database Lets You See How Celebrities Tip' (2014) Gawker 23 October http://gawker.com/the-public-nyc-taxicab-database-that-accidentally-track-1646724546 (accessed 25 August 2017).

of 3 months of credit card transactions, chronicling the spending of 1.1 million people in 10,000 shops, having access only to amounts spent, shop type and a code representing each person. Knowing the amounts spent on these four occasions led to re-identification of nearly all card-holders.[43]

The examples demonstrate that the capacity of (re-)identification is increasing each year, and although perfect identification may still be a myth today, one wonders for how long. Many legal (eg the often cited Ohm[44]) and technical scholars (eg Narayanan[45]) agree that at this rate, a meaningful distinction between identifiable and non-identifiable information is not sustainable much longer.[46] In Europe this conclusion is further reinforced by the complementary WP29 opinion on anonymisation: the data is not identifiable, ie anonymous, only when anonymisation is irreversible.[47]

3.3. 'Any information'

Unlike 'identifiability', the remaining two elements of the definition receive little attention in the literature. Yet, as the WP29 interprets them, they significantly contribute to the imminent explosion of the kinds of data that can be considered personal, and are therefore not less controversial. While explaining the meaning of 'any information', the WP29 does not examine what information means, probably considering it self-evident, and focuses immediately on *what kinds of information* would fall under 'any information'. In the age of ubiquitous computing and advanced analytics algorithms, this omission may have a truly explosive effect on the range of situations that would fall within the scope of data protection law.

3.3.1. Broad but undefined concept of information

WP29 starts with a broad declaration of the intent of the legislator 'to design a broad concept of personal data'[48] and further explains that *any* information can fall under the concept 'personal data' regardless of its nature, content, or format. To be considered personal data, the *nature* of information is of no significance: it can be true or inaccurate, objective and subjective, including opinions and assessments.[49]

[43] J Bohannon, 'Credit Card Study Blows Holes in Anonymity' (2015) 347(6221) *Science Magazine* 468.
[44] Ohm (n 6) 1701.
[45] A Narayanan and EW Felten, 'No Silver Bullet: De-Identification Still Doesn't Work' (2014) randomwalker.info, published 9 July 2014 http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf (accessed 18 February 2018).
[46] But see A Cavoukian and D Castro, 'Big Data and Innovation, Setting the Record Straight: De-identification Does Work' (2014), www2.itif.org/2014-big-data-deidentification.pdf (accessed 18 February 2018).
[47] Article 29 Working Party opinion 05/2014 on Anonymisation Techniques, 10 April 2014 (WP 216), 3, 5–7.
[48] WP 136 (n 15), 6.
[49] Ibid, 6.

Next, there are no particular requirements to the *content* of information. Information does not have to concern private or family life, and could pertain to the life of the individual in his or her professional and other capacities,[50] which is consistent with the aim of the DPD to protect 'the fundamental rights and freedoms of natural persons, and in particular [but not exclusively] their right to privacy'.[51]

Finally, information can constitute personal data regardless of the *format*, medium, or form, which could be 'alphabetical, numerical, graphical, photographical or acoustic', 'kept on paper [or] stored in a computer memory' as a binary code,[52] structured or unstructured,[53] provided the other criteria of the definition are met. Video and voice recording can be such information, as well as a child's drawing that could contain personal data of both the child and the parents.[54]

Aside from this, the explanation by WP29 refers to 'information' as a concept the meaning of which is self-evident. In any case, no further clarification as to what is meant by 'information', and how it relates to or differs from other information-related concepts, such as 'data', 'meaning', 'knowledge', or information artefacts ('information carriers', eg books, CDs, etc.), is given.[55] The only exception is a short paragraph regarding human tissue samples. According to WP29, these are 'sources' of biometric data but 'not biometric data themselves.'[56]

> Therefore the extraction of information from the samples is collection of personal data, to which the rules of the Directive apply. The collection, storage and use of tissue samples themselves may be subject to separate sets of rules.[57]

This paragraph suggests that the samples are information carriers, like eg CDs, rather than information. At the same time, the exact wording of the paragraph is not this explicit. Specifically, a number of questions arise: first, is there a meaningful difference between tissue samples and the child's drawing that the WP29 used earlier as an example of information and not a source of information, when both do not provide information immediately but require 'information extraction'; second, if the tissue samples 'are not biometric data themselves', are they biometric data in combination with

[50]Ibid, 6–7.
[51]Ibid, 7.
[52]Ibid.
[53]Ibid, 8.
[54]Ibid.
[55]Bygrave considers it problematic that the legal definition of personal data is 'creeping' from the intangible world onto the world of biomaterial (L Bygrave, 'Information Concepts in Law: Generic Dreams and Definitional Daylight' (2015) 35(1) *Oxford Journal of Legal Studies* 91–120); WG Urgessa, 'The Feasibility of Applying EU Data Protection Law to Biological Materials: Challenging 'Data' as Exclusively Informational' (2016) 7 *JIPITEC* 96.
[56]WP 136 (n 15), 9.
[57]Ibid.

additional information or in a certain technological context; and third, if the collection, storage and use of such samples 'may be subject to separate sets of rules', are those rules in addition to data protection, *lex specialis* in relation to the general rules of the DPD, or does the data protection law not apply?

3.3.2. 'Everything is information'

The striking consequence of the declared commitment of WP29 to the all-encompassing interpretation of 'any information' and the vagueness of the WP136 on the meaning of information is that a viable argument can be built that *everything is or at least contains information*.

Information is a notoriously nebulous concept that has different meanings. These meanings vary over time, across and within disciplines as diverse as philosophy, psychology and cybernetics, or depending on one's philosophical inclination,[58] resulting in what Burgin calls 'information studies perplexity'.[59]

To use Burgin's illustration, a popular definition of information ('information is the eliminated uncertainty'[60]) is based on Claude Shannon's information theory,[61] which represents a statistical approach to information.[62] This theory and the information definitions based on it are criticised for completely 'ignoring the human aspect of information' and misleading social sciences and humanities.[63] Yet, Claude Shannon never intended to develop a theory of information (it was originally called 'communication theory' and was relabelled by his followers later) or define what information is.[64] Some scholars define information through knowledge or data, and others define knowledge and data in terms of information.[65]

While law is generally characterised by poor conceptualisation of information,[66] several analyses have adopted a General Definition of Information ('GDI') as an operational standard: '*information is data + meaning*'.[67] Data is the first element of the definition and stands for the lack of uniformity in the

[58] L Floridi, 'Philosophical Conceptions of Information' in Giovanni Sommaruga (ed), *Formal Theories of Information: From Shannon to Semantic Information Theory and General Concepts of Information* (Springer, 2009), 13–53, 16; P Adriaans, 'Information', *The Stanford Encyclopedia of Philosophy* (2013) http://plato.stanford.edu/archives/fall2013/entries/information (accessed 18 February 2018).

[59] M Burgin, *Theory of Information. Fundamentality, Diversity and Unification* (World Scientific Publishing, 2010) 6.

[60] Ibid, 6, referring to RV Hartley, 'Transmission of information' (1928) 7 *The Bell System Technical Journal* 335–363 and AD Ursul, *Information* (Nauka, 1971, in Russian).

[61] CE Shannon, 'A mathematical theory of communication' (1948) 27 *The Bell System Technical Journal* 379–423.

[62] Burgin (n 59) 6.

[63] Ibid, 6, citing L Brillouin, *Science and Information Theory* (Academic Press, 1956).

[64] Burgin (n 59) 6.

[65] Ibid, 7 and in-text references therein on knowledge, and 192 on the relationship between information and data (eg 'information is data + meaning' vs 'data is potential information').

[66] For an elaborate discussion of the concept of information in law see Bygrave (n 55) 91–120; examples of legal arguments against defining information in law in Bygrave (n 55), footnotes 5 and 6.

[67] L Floridi, 'Is Information Meaningful Data?' (2005) 70(2) *Philosophy and Phenomenological Research* 351–370.

world (what Floridi poetically calls 'the fractures in the fabric of being'[68]), or between at least two physical states (a higher or lower charge in a battery), or between two symbols (letters A and B, or numerals 0 and 1).[69] Put differently, data is 'a description of something that allows it to be recorded, analysed, and reorganized'.[70] Meaning is the second element of the GDI. To perceive data as information we need to make sense of it. For instance, we understand combinations of letters as words and sentences, or observe that the battery charge is low and conclude that the battery will soon need to be recharged.

While the definition of information as data + meaning has been broadly adopted,[71] others disagree and reject the idea that information is always meaningful to those who use it. As Hildebrandt explains,[72]

> Though many organisms do not speak our type of language, they all depend on information to survive and flourish. Artificial intelligence similarly depends on the processing of information (in the form of digital data points). Neither in the case of organisms nor in the case of intelligent machines does information necessarily imply the attribution of meaning, as it may in the case of humans.[73]

In other words, meaning is a function of 'the curious entanglement of self-reflection, rational discourse and emotional awareness' that is (still) a human prerogative, and cannot be a defining element of information which non-humans (animals and machines) process, too.[74] This is a part of a long-standing debate about whether or not information belongs exclusively to the domain of human society, or is to be found everywhere in the universe.[75] It appears that a growing number of physicists 'define the physical world as being made of information itself' and some even argue that the so-called 'structural and kinetic information is an intrinsic component of the universe […] independent of whether any form of intelligence can perceive it or not'.[76]

It is beyond the ambition of this paper to jump down the rabbit hole of the various definitions of information, meaning and data, in search of the most correct ones. It suffices to point out that adopting a broad approach to information as the WP29 does leaves the concept of personal data wide open to potentially apply to literally everything, provided other conditions are met.

[68]Floridi (n 58) 17.
[69]Ibid.
[70]V Mayer-Schönberger and K Cukier, *Big Data: A revolution That Will Transform How We Live, Work, and Think* (Houghton Mifflin Harcourt, 2009), Chapter 5 (e-book).
[71]Burgin (n 59) 188 et seq brings a number of examples.
[72]M Hildebrandt, 'Law As Computation in the Era of Artificial Legal Intelligence. Speaking Law to the Power of Statistics' (forthcoming) *University of Toronto Law Journal* 10, also see references in fn 48 https://ssrn.com/abstract=2983045 (accessed 18 February 2018).
[73]Ibid.
[74]Ibid, 10.
[75]Burgin (n 59) 33.
[76]Ibid, and in-text references.

If we adopt the broadest understanding of information as adopted eg by Hildebrandt, ie not dependent on human thinking, then 'the whole nature is a huge system of information processes', what Wheeler, a prominent physicist, calls 'from It to Bit',[77] where 'each physical situation […] emerges from the flow of information, generates information and gives information to the environment'.[78] In other words, everything is information; it is impossible to distinguish between a tree and information about a tree for a tree is made of information.

If we adopt a narrower understanding of information dependent on humans or other recipients attributing meaning or significance to data, ie the so-called semantic concept of information epitomised in the GDI, the following argument unfolds.[79] Information is data + meaning or significance. People measured and quantified the world long before the Digital Age, eg in the form of maths and language. However, the arrival and proliferation of computing turbocharged this process, enabling the measurement and quantification of, and hence the harnessing of data from, literally everything.[80] Hence, everything can be a source of data. But to be information, does all data have meaning or significance? This depends on the recipient. As Burgin explains, the same information can have no or various meanings depending on who perceives it. For instance a text in Chinese only has meaning to a speaker of Chinese, and a paper on mathematics has no or different meaning to a lay reader, high-level mathematician, or a mathematics major.[81] Human mentality[82] is seen by some (eg Hildebrandt) as the precondition for a meaning to occur, while others see computers having mentality as well in the form of everything in the computer's memory.[83] Human mentality is limited in the extent to which it can make sense of data. Some data colloquially referred to as 'raw' has no or very limited meaning in human perception, like zeros and ones of a binary code. The way all *computers attach meaning to data* is *different from human cognition*, eg computers have a different language and process data faster. Humans – albeit with proper

[77] JA Wheeler, 'Information, Physics, Quantum: The Search for Links', in WH Zurek (ed), *Complexity, Entropy, and the Physics of information* (Addison-Wesley, 1989), 3–28.

[78] Burgin (n 59) 34 citing T Stonier, *Information and Meaning: An Evolutionary Perspective* (Springer, 1997).

[79] The semantic concept of information is heavily criticised for a presumption that data is meaningless: '[T]here are no "raw" data as any observable, measurable and collectible fact has been affected by the very knowledge that made this fact observable, measurable and collectible' (Burgin (n 59) 197).

[80] Mayer-Schönberger and Cukier (n 70).

[81] Burgin (n 59) 9–10, 94, citing CT Meadow and W Yuan, 'Measuring the Impact of Information: Defining the Concepts' (1997) 33(6) *Information Processing and Management* 697–714 and Fred Dretske, *Knowledge and the Flow of Information* (Basil Blackwell, 1981).

[82] Mentality is a part of the so-called 'Existential Triad of the world' which describes the world's structure: physical world, mental world and the world of structures (Burgin (n 59) 60). The Existential Triad is based on Popper's triad of the world: physical objects or states, consciousness or psychical states, and intellectual contents of books, documents, scientific theories, etc. (KR Popper, 'Replies to my critics', in PA Schilpp (ed), *The Philosophy of Karl Popper* (Open Court, 1974) 949–1180.

[83] Burgin (n 59) 61.

training and effort – can still follow how traditional computers make sense of data: these computers follow deductive models, rules and cases.[84] However, how meaning is 'attached' to data by modern machines is beyond the grasp of human mind.[85] The game-changer is a new generation of data-processing algorithms based on machine learning. Machine learning is the ability of computer algorithms to learn from data and make predictions for new situations,[86] and improve automatically through experience.[87] The new algorithms are autonomous, ie self-learning, self-repairing, and self-managing, and form the core of the modern approach to Artificial Intelligence ('AI'), a strand of computer science aiming to build computers as intelligent agents. The way advanced AI self-learning algorithms make sense of data is not transparent even for their designers. Hence, the new AI algorithms work as a *black box* that is truly beyond human cognition.[88] These AI self-learning autonomous machines together with unprecedented amount of data already stored in databases or live-streamed form the essence of Big Data[89] and have the ability to harness information in fundamentally novel ways.[90] In effect, we can no longer say that some data has no meaning for we really have lost to computers the monopoly of deciding that. In fact, it is safer to assume that all data – the number and frequency of steps or key strokes one makes daily, the colour of one's eyes or even how many leaves grow on a tree – potentially has meaning, even if not for humans. Hence, everything is data and all data has meaning; hence, everything is or contains information.

3.4. 'Relating to'

What 'relating to' means, according to WP29, is 'crucial as it is very important to precisely find out which are the relations/links [between the information and the individual] that matter and how to distinguish them'.[91] Again, the WP29 construes 'relating to' very broadly.

WP29 points out that in some situations this relationship is quite obvious, while in others the link is not self-evident. In particular, the latter is the case

[84]M Hildebrandt, 'Slaves to Big Data. Or Are We?' (2013) 16 *IDP Revista De Internet, Derecho Y Política* http://works.bepress.com/mireille_hildebrandt/52/ (accessed 10 October 2017).
[85]M Hildebrandt, 'The Dawn of a Critical Transparency Right for the Profiling Era' in J. Bus et al (eds), *Digital Enlightenment Yearbook* (IOS Press, 2012), 53; JP Van Bendegem, 'Neat Algorithms in Messy Environments' in M Hildebrandt and S Gutwirth (eds), *Profiling the European Citizen. Cross-Disciplinary Perspectives* (Springer, 2008), 80–83.
[86]Jelena Stajic, Richard Stone, Gilbert Chin, and Brad Wible, 'Rise of the Machines' (2015) 349(6245) *Science* 248–249.
[87]MI Jordan and TM Mitchell, 'Machine Learning: Trends, Perspectives, and Prospects' (2015) 349(6245) *Science* 255.
[88]Hildebrandt (n 85), Van Bendegem (n 85).
[89]Hildebrandt (n 84).
[90]Mayer-Schönberger and Cukier (n 70).
[91]WP 136 (n 15), 9.

where information relates to an object, eg the value of a house, or a process or event, eg data on the functioning of a machine that requires human intervention. In these cases there would be an indirect relationship to people owning or otherwise interacting with the object. Presumably, WP29 considers the indirect relationship sufficient. In all cases, all circumstances of the case need to be taken into account in assessment.[92]

Information can 'relate' to an individual in *content*, *purpose*, or *result*, meaning that information 'relating to' a natural person includes but is broader than the information 'about' that person. The meaning of 'relating to' grows even broader considering that these three conditions are meant as alternative and not as cumulative ones.[93]

Information relates to a person more obviously in content, ie when it is *about* that person. However, even the information that is not in any way *about* someone may be found to 'relate to' a person. Information relates to a person in purpose 'when the data *are used or are likely to be used … with the purpose* to evaluate, treat in a certain way or influence the status or behaviour of an individual'[94] [emphasis added]. Finally, information regardless of its content or any purpose of processing may relate to a person in result when '[its] use is *likely to have an impact* on a certain person's rights and interests'.[95] According to the WP29, the relevant impact does not have to be 'major',[96] which I understand to mean that the information relates to a person by reason of impact even if it is likely to affect that person in only a minor way. WP29 further clarifies that '[i]t is sufficient if the individual may be treated differently from other persons as a result of the processing of such data'.[97]

Remarkably, similar to the criterion of identifiability, the relationship by reason of purpose and result will occur not only when data *is already used*, but also where it is *likely to be used* with the purpose or effect of impacting people 'taking into account all the circumstances surrounding the precise case'.[98] I have already discussed how the reference to likelihood results in the standard of the reasonable likelihood of identification that is both broad and context-dependent, making the status of data as 'personal' dynamic (see 3.2). The same argument equally applies here, with a significant correction: while the likelihood of identification under Recital 26 must be 'reasonable', which WP29 interprets as more than a purely hypothetical possibility, there is no such clarification here, and hence a lower threshold of what is likely to relate to a person applies.

[92] Ibid, 9–10.
[93] Ibid, 11.
[94] Ibid, 10.
[95] Ibid, 11.
[96] Ibid.
[97] Ibid.
[98] Ibid.

In other words, some information is perceived as relevant more easily, for instance, information 'generated' *by* (observing) people (eg administrative records of people's offline lives, and digital records of online behaviour such as websites visited, texts and images uploaded; information generated through use of 'smart' objects and devices such as phones or fitness bracelets), or objects people interact with (their cars, homes, computers). At the same time, some information is hard to intuitively place in any connection of relevance for anyone: eg the amount of weight a block of concrete can withstand, or the number of sand crystals in a cubic metre of sand in the Sahara desert. However, when increasing amounts of data are gathered in real time from increasingly connected environments, intended to be used in automated decision-making about us, and we do not know how the autonomous self-learning and self-managing computers draw meaning from data, we should always reasonably assume that any information is likely to relate to a person, since we cannot eliminate this possibility with certainty. The latter point requires further elaboration.

Not all information relates to people by reason of its content, eg is of biographical significance, describes a person or conveys a fact from someone's life or assessment of a person. However, in the emerging world of data-driven decision-making, cyber-physical infrastructures and what Hildebrandt calls 'data-driven agency',[99] any information can relate to a person by reason of purpose, and all information relates to a person by reason of impact.

The relationship by reason of a purpose, as the WP29 explains it, occurs 'when the data *are used or are likely to be used … with the purpose* to evaluate, treat in a certain way or influence the status or behaviour of an individual'.[100] This essentially describes the relationship of information to a person in terms of intended impact. By now much of the information is processed specifically with the intent to impact people in the way the WP29 describes, ie 'to evaluate, treat in a certain way or influence the status or behaviour'. For instance, to treat in a certain way or influence human behaviour (eg trigger people to buy a certain product or conserve energy) is the chief reason for information collection and further processing on the ad-powered Internet and in the context of the data-driven (or algorithmic) regulation.[101] The ultimate destination, however, is the world of data-driven agency, which Hildebrandt defines as 'a specific type of artificial intelligence, capable of perceiving an

[99]'Data-driven agency refers to a specific type of artificial intelligence, capable of perceiving an environment and acting upon it, based on the processing of massive amounts of digital data' (M Hildebrandt 'Law as Information in the Era of Data-Driven Agency' (2016) 79(1) *MLR* 4).

[100]WP 136 (n 15), 10.

[101]Yeung defines algorithmic regulation through algorithmic decision-making: 'Algorithmic decisionmaking refers to the use of algorithmically generated knowledge systems to execute or inform decisions, which can vary widely in simplicity and sophistication. Algorithmic regulation refers to regulatory governance systems that utilize algorithmic decisionmaking.' K Yeung, 'Algorithmic Regulation: A Critical Interrogation' (forthcoming) *Regulation & Governance* doi:10.1111/rego.12158.

environment and acting upon it, based on the processing of massive amounts of digital data'.[102] A 'smart' city where all aspects of the environment and people living in it are datified, and the inhabitants are subjected to a certain treatment in real time based on processing of the data, from the speed at which escalators are running to promote physical activity to the warmth and intensity of street lighting to prevent undesirable behaviour to targeted policing, would be an example of such data-driven agency. As our homes and cities are increasingly being made 'smart', laced with the network of communicating sensors built into the infrastructure, mobile phones, toys and appliances, and as information technology is increasingly mediating our daily functioning, the onlife world of data-driven agency materialises more clearly. At present, the 'narratives of a frictionless world that surreptitiously adjusts the environment to the needs and desires of its users'[103] are steadily on the way out of the realm of science fiction.[104] In such a world, any information within the 'smart' environment can be used and all information is likely to be used with the purpose of adapting the environment and impacting people.

Finally, information regardless of its content or purpose of processing may relate to a person in result when '[its] use is *likely to have an impact* on a certain person's rights and interests' even in a minor way, eg different treatment.[105] This describes the relationship of information to a person in terms of the impact that is inadvertent rather than intended. The hyper-connected world of the data-driven agency described above provides fertile ground for such accidental impact. Indeed, unpredictability of outcomes is considered one of the characteristic features of the advanced data analytics running at the back-end of this world.[106] This is also why some argue that Big Data analytics powered by machine learning is at odds with the purpose limitation principle of data protection. Hence, all information in the context of the data-driven agency relates to people by reason of impact.

Hildebrandt has recently questioned the assumption of unpredictability of machine learning. She argues that 'the methodological integrity of the machine learning requires advance specification of the purpose as this will

[102]Hildebrandt (n 99), 4.
[103]Ibid.
[104]In addition to fully adaptive digital environments, the vision of fully adaptive physical environments is also already here. Eg Maas et al sketch a fictional scenario of the cities of the future built of a fictional Barba material that can be programmed to easily transform itself in real time to meet our desires and needs, eg for more or different space. The book explores how modern developments in robotics, material science and computing can enable this scenario and supports the vision with programming experiments and applied prototypes. The book is based on a series of projects where the Delft University of Technology, ETH Zürich and the European Institute of Innovation and Technology collaborated (W Maas et al, *Barba. Life in the Fully Adaptable Environment* (nai010, The Why Factory 2015).
[105]WP 136 (n 15), 11.
[106]See eg T Zarsky, 'Incompatible: The GDPR in the Age of Big Data' (2017) 47 *Seton Hall Law Rev*iew 995, 1005 ('Big Data involves methods and usage patterns which neither the entity collecting the data nor the data subject considered or even imagined at the time of collection.').

inform the solidity and productivity of the relevant research design'.[107] If this is the case, it should be possible to distinguish between the information that will be likely to impact people, and the information that will not. Specifying the intended outcome of data analytics may be good science or, in Hildebrandt's words, an 'agonistic' approach to machine learning.[108] Yet, this does not negate my argument until advance purpose specification and discarding accidental outcomes become practice. In fact, the almost religious belief in the unpredictability and nearly magic powers of the Big Data to see previously unseen patterns and 'generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' forms what Boyd and Crawford call 'mythology' of Big Data which has become a constitutive element of the phenomenon 'Big Data' next to the large data sets, computational power and advanced algorithms.[109] All information is likely to have an impact on people as long as this mythology has a tangible effect on how data analytics is used.

3.5. Playing devil's advocate: weather is personal data

While the preceding analysis covered the elements of the definition of personal data according to the WP29 and looked into the future of the concept in the emerging onlife world, this section will focus on the present. I will demonstrate with the use of a provocative example how information which intuitively is far from being 'personal' can be plausibly argued to fit the definition and hence is 'personal data'.

Weather, while neutral and of no consequence as a subject of small talk, may be considered information relating to an identified or identifiable natural person in the context of a living lab, for instance the Stratumseind 2.0 smart city project in Eindhoven, the Netherlands.[110] In short, Stratumseind is the longest 'going-out' street in the Netherlands, with many bars and 'coffeeshops'. The area struggles with the declining numbers of visitors and a decrease in turnover of the local businesses. Criminal behaviour and vandalism on the street are seen to cause the problem. In response, the municipality of Eindhoven, together with the police and a number of private parties (the Stratumseind establishments association, real estate owners, etc.), initiated a smart city project *Stratumseind 2.0*. The project includes a number of smaller data-driven projects, some of which aim at predicting,

[107]M Hildebrandt, 'Privacy as Protection of the Incomputable Self: Agonistic Machine Learning' (forthcoming) *Theoretical Inquiries in Law*, available online https://ssrn.com/abstract=3081776, 13 (SSRN page numbering).
[108]Ibid.
[109]D Boyd and K Crawford, 'Critical Questions for Big Data. Provocations for a Cultural, Technological, and Scholarly Phenomenon' (2012) 15(5) *Information, Communication & Society* 663.
[110]In the factual description of Stratumseind 2.0 I refer to the ongoing PhD research of Maša Galic on privacy and big data-based security in public spaces.

preventing and de-escalating deviant behaviour on the street, among other things, by engaging the police or adapting the street lighting. Private companies such as Atos, Intel and Philips see Stratumseind as a test-bed for their smart-city products and supply the technology enabling the projects.

The data is gathered from the multiple sensors installed on the street, including video- and acoustic cameras, sound sensors, WiFi tracking and a weather station. These sensors, among others, measure how many people pass by the street per day, how they move, where they come from, sound on the street, what it is and where it comes from, rainfall per hour, temperature, wind direction and speed. All the data is stored in a database which enables knowledge discovery, although the actual currently performed analysis is limited.

Would rainfall per hour, temperature, wind direction and speed, together referred to as 'weather', be personal data following the WP29 guidelines? I would argue that they are. The weather is information if we adopt an approach to information as an omnipresent and all-encompassing phenomenon, captured by Wheeler's 'from It to Bit'.[111] Alternatively, according to the semantic definition of information, weather contains information because it is observed and recorded by the weather station, ie 'datified'. Both interpretations fit the broad approach of the WP29. Although not *about* people, this information is collected in a database that is likely to be used for a *purpose* to assess and influence their (deviant) behaviour, and hence it is information *relating to* people in purpose. In the context of a large knowledge-discovery database built for advanced data analytics, by several different public and private parties with varying interests, it is possible to imagine that not all results of data analytics pertaining to human behaviour on the street would be intended, in terms of both the opaqueness and unpredictability of the advanced analytics algorithms and the possible lack of absolute agreement between the involved parties on their intentions. In that aspect, the weather information will still be relating to people in *impact*. Finally, each visitor to Stratumseind is highly likely to be identifiable if not by the weather information alone, certainly in combination with the data from the WiFi tracking sensors, voice recordings or video footage; if not by a weather station operator, certainly by some other project partners who, being technology companies, possess the tools and expertise to do so. Alternatively, since the purpose of the project is not only to predict but also to address criminal and other deviant behaviour, this implies that the perpetrators' *identification is intended*. Hence, the identification is '*reasonably likely*', as the WP29 explains. Therefore, under these circumstances, depending on the approach to information, either weather itself is information or information about weather is information that relates to a number of personally identifiable

[111]Wheeler (n 77) 3–28.

natural persons, and is personal data. A similar argument may be constructed in relation to waste water, should the project partners decide to include the data from monitoring the sewer (eg for the purpose of detecting drug labs) in the database.

An objection may be raised that the weather and waste water could only be brought in relationship to an identifiable person and hence be or contain personal data in very particular circumstances of what Cavoukian and Castro call 'high-dimensional data', which 'consists of numerous data points about each individual, enough that every individual's record is likely to be unique, and not even similar to other records'.[112] However, as Narayanan and Felten explain, 'high-dimensional data is now the norm, not the exception. … [T]hese days it is rare for useful, interesting datasets to be low-dimensional'.[113]

This weather example may come across as provocative, but illustrates a real pattern, a general direction in which the concept of personal data is developing. More and more situations will constitute processing of personal data as interpreted by the WP29 until the onlife world will eventually stop being a scenario of the (near) future and become reality. This will eventually turn data protection law into an uneconomical exercise of *regulating everything*, and deprive its protection of meaning. As a result, the current personal data-centred paradigm of legal protection will not be sustainable in the long run.[114]

4. The Court of Justice case law

4.1. The significance of the WP29 opinions for the Court of Justice's jurisprudence

This part will examine whether the trend of data protection law turning into the law of everything is materialising in the case law of the EU Court of Justice, ie if the Court's interpretation of 'personal data' is in line with WP29.

As explained earlier, while the WP29 opinions are highly influential in data protection practice and are key compliance guidelines, they are not a source of data protection law and not binding. Under Article 29(1) DPD, the role of the Working Party is 'advisory'. Therefore, when ruling on the meaning of personal data, the Court is free to set aside WP29 opinions, including the WP136. As a matter of fact, the Court never cites WP29 opinions in its data protection jurisprudence,[115] and on a number of occasions has ruled contrary to the WP29's earlier opinions. A recent example is the *Google*

[112]Narayanan and Felten (n 45).
[113]Ibid.
[114]Part 5 explains why data protection law as 'the law of everything' is problematic.

Spain case, where the Court found search engine providers to be controllers with regard to indexing and making available via its search personal data published on third-party websites.[116] Hereby, without expressly mentioning it, the Court overruled the earlier position adopted by the WP29 on two points: first, that the search engine providers cannot be 'the principal controller' when they act 'purely as an intermediary' with regard to the third-party content containing personal data, and second, that the search engine providers can only be controllers 'with regard to the removal of personal data from their index and search results, [while] the extent to which an obligation to remove or block personal data exists, may depend on the general tort law and liability regulations of the particular Member State'.[117]

At the same time, Advocates General ('AGs') in their opinions for the Court do, although not always, refer to the WP29.[118] Hence the WP136 may still be of indirect influence on how the concept 'personal data' develops in the case law, if not in terms of substantive outcomes, surely in terms of providing the AGs and the Court with a list of issues to consider.

4.2. The development of the Court of Justice's case law on the concept of personal data

This section will very briefly sketch the development of the EU case law on the meaning of personal data from *Lindqvist* until roughly 2014, focusing on the big lines rather than exhaustive description. The year 2014 is chosen as a milestone as it marks a different stage in the Court of Justice case law. This was the year when the Court for the first time engaged in a discussion of an element of the definition of personal data and produced extensive analysis of the meaning of 'information relating to' a person in *YS and others*.

The Court of Justice ruled on the meaning of personal data in the Directive in its very first data protection case, *Lindqvist*,[119] and on a number of occasions since then. However, the Court's judgments are not nearly as comprehensive as the WP136. Part of the reason is that nearly all of them are given in the context of a reference for preliminary ruling. While it is true that the Court often reformulates the questions asked, it remains constrained by the questions from the national courts and the circumstances of each case.

[115]A search done on 28 August 2017 for 'Article 29 Working Party' using the http://curia.europa.eu search tool, under 'data protection' subject matter, resulted in four hits, all opinions of Advocates General and no judgments of the Court.

[116]Case C-131/12 Google Spain SL, Google Inc v Agencia Española de Protección de Datos and Mario Costeja González [2014] ECLI:EU:C:2014:317 [31] et seq.

[117]Article 29 Working Party Opinion 1/2008 on data protection issues related to search engines, adopted on 4 April 2008 (WP 148), 14.

[118]Most recently *Breyer* (n 11), Opinion of Advocate General Sanchez-Bordona; and *Nowak* (n 24), Opinion of Advocate General Kokott.

[119]Case C-101/01 *Bodil Lindqvist* [2003] ECR I-12992.

In its case law the Court often states that the scope of the Directive is very wide and that the personal data covered by the Directive is varied.[120] Most of the relevant cases simply name a particular type of data involved, ruling that this information indeed constitutes personal data. There is no discussion of what elements the concept of personal data entails and what each of those elements should mean. In *YS and others*, AG Sharpston[121] gives some examples of such types of data which the Court has explicitly pronounced personal: 'the name of a person in conjunction with his telephone coordinates or information about his working conditions or hobbies',[122] his address,[123] his daily work periods, rest periods and corresponding breaks and intervals,[124] monies paid by certain bodies and the recipients,[125] amounts of earned or unearned incomes and assets of natural persons.[126]

Only relatively recently, as the data processing situations which the national courts deal with have become more complex, have the courts begun to refer questions that require a more detailed analysis of what particular elements of the concept 'personal data' mean. The subsequent sections will analyse the EUCJ's case law that resulted, specifically, *Breyer*, *YS and others* and *Nowak*. The analysis will show that the case law of the Court of Justice is either in line with the broad approach to the notion of personal data that the WP29 adopts, or has limited potential to avert the imminent explosion of the situations falling within the scope of data protection law.

4.3. Case law on the scope of data protection and unreasonable and disproportionate legal consequences

As the notion of personal data largely determines the material scope of data protection law, construing the elements of personal data broadly inevitably results in delineating the scope of data protection law widely. The Court has dealt with the argument that interpreting the scope of the DPD too broadly would lead to outcomes which are unreasonable and disproportionate. For instance, in *Lindqvist*, the argument on behalf of Mrs Lindqvist was that it was 'unreasonable to take the view that the mere mention by name of a person or of personal data in a document … on an internet page constitutes

[120]Joined Cases C-465/00, C-138/01 and C-139/01 *Österreichischer Rundfunk and Others* [2003] ECR I-4989 [43]; *Lindqvist* (n 119) [88]; and Case C-553/07 *College van burgemeester en wethouders van Rotterdam v M.E.E. Rijkeboer* [2009] ECR I-3889, [59].
[121]Joint cases C-141/12 and C- 372/12 *YS and M. and S. v Minister of Immigration, Integration and Asylum* [2016], ECLI:EU:C:2014:2081, Opinion of Advocate General Sharpston [44].
[122]*Lindqvist* (n 119) [24].
[123]*Rijkeboer* (n 120) [62].
[124]Case C-342/12 *Worten Equipamentos para o Lar SA v Autoridade para as Condições de Trabalho (ACT)* [2013] OJ C225/37, [19], [22].
[125]*Österreichischer Rundfunk and Others* (n 120) [64].
[126]Case C-73/07 *Satakunnan Markkinapörssi and Satamedia* [2008] ECR I-09831, [35], [37].

automatic processing of data'.[127] In the more recent *Google Spain* case, AG Jääskinen submitted that

> the broad definitions of personal data, processing of personal data and controller are likely to cover an unprecedently wide range of new factual situations … . This obliges the Court to apply […] the principle of proportionality, in interpreting the scope of the Directive in order to avoid unreasonable and excessive legal consequences.[128]

However, in both cases, the Court did not accept that the proportionality and 'reasonableness' had to be taken into account at the stage of determining the Directive's scope. Rather, according to *Lindqvist*, the Directive itself has a degree of flexibility[129] to enable proportionate application of its rules. In effect, the principle of proportionality is respected at the stage of national implementation of the Directive and is secondary to the issue of scope.[130] Similarly, in *Google Spain* the Court did not accept the proportionality argument, given the aim of the Directive 'to ensure a high level of protection of the fundamental rights and freedoms … with respect to the processing of personal data',[131] and that in so far as the Directive's provisions are 'liable to infringe fundamental freedoms', they must be interpreted in the light of fundamental rights.[132] Hence, the position of the Court so far has been that any possible undesirable impact of the broad application of data protection law should be mitigated not through narrower interpretation of the scope, but rather through proportionate application of particular data protection provisions. The case law on the individual elements of the concept 'personal data' largely fits within the same pattern.

4.4. 'Identified or identifiable': Breyer

The Court dealt with the meaning of 'identifiable' in the reference for a preliminary ruling in the *Breyer* case. The reference was brought in 2014 and the ruling was given in 2016. The judgment was generally received as a confirmation of the absolute approach to identifiability in EU data protection law which was first declared in Recital 26 DPD, then adhered to by the WP29, and finally received the assent of the Court.[133] I will argue, however, that while the Court indeed went for a broad interpretation of identifiability, its

[127] *Lindqvist* (n 119) [20].
[128] *Google Spain* (n 116), Opinion of Advocate General Jääskinen [30].
[129] *Lindqvist* (n 119) [83].
[130] *Lindqvist* (n 119), [87]–[88].
[131] *Google Spain* (n 116), [66].
[132] Ibid, [68]. For a general discussion of the judgment see O Lynskey, 'Control over Personal Data in a Digital Age: Google Spain v AEPD and Mario Costeja Gonzalez' (2015) 78(3) *Modern Law Review* 522.
[133] Eg FJ Zuiderveen Borgesius, 'Breyer Case of the Court of Justice of the European Union: IP Addresses and the Personal Data Definition' (2017) 3(1) *European Data Protection Law Review* 130.

reading is more restrained and measured compared to the WP136, and narrows down the scope of the concept 'personal data' somewhat.

In *Breyer*, the main dispute concerned Mr Breyer, a German activist, and the Federal Republic of Germany. The former visited some publicly accessible websites of the German Federal institutions, and the latter, for the purposes of preventing denial of service attacks, stored IP addresses of the websites' visitors, including Mr Breyer's dynamic IP addresses. Mr Breyer objected to the retention of his dynamic IP addresses as his personal data.

On an earlier occasion, in *Scarlet Extended*, the Court had already found static IP addresses to be personal data for the Internet service provider 'because they allow users to be precisely identified' by the Internet service provider.[134] However, *Breyer* presented a different issue. Namely, unlike a static IP address 'which [is] invariable and allow[s] continuous identification',[135] a dynamic IP address changes each time there is a new Internet connection, and does not allow a link to be established, 'through files accessible to the public', between a given computer and the Internet connection.[136] In fact, the Federal Republic of Germany, the website provider who kept the IP addresses, could not identify Mr Breyer through his dynamic IP address[137] and would need additional information in possession of the Internet service provider to be able to do so. It was 'common ground' that a dynamic IP address does not constitute information relating to an *identified* natural person, since the identity of the owner or another user of a computer bearing the IP address is not directly revealed.[138] Hence, the central issue before the Court was whether such an IP address constitutes information relating to an *identifiable* natural person in relation to the website provider where the additional data necessary for identification of the website visitor was held by the visitor's Internet service provider.[139]

This time, the Court skipped the usual reference to the scope of the Directive, which is very wide, and the personal data covered by the Directive are varied.[140] The argument developed in several steps: first, to be considered personal data, it is not necessary for information on its own to allow identification of the data subject (this follows from the definition of personal data as relating to an identifiable person, either directly or *indirectly*).[141] Second, neither is it necessary 'that all the information enabling the identification … must be in the hands of one person',[142] which follows from Recital 26

[134] *Breyer* (n 11), [33]–[34], citing *Scarlet Extended* (C-70/10, EU:C:2011:771), [51].
[135] Ibid, [36].
[136] Ibid, [16].
[137] Ibid, [35], [37].
[138] Ibid, [38].
[139] Ibid, [39] (emphasis added).
[140] *Österreichischer Rundfunk and Others* (n 120) [43]; *Lindqvist* (n 119) [88]; and *Rijkeboer* (n 120) [59].
[141] *Breyer* (n 11), [40]–[41].
[142] Ibid, [43].

DPD and its two-fold standard of (1) 'all the means likely reasonably to be used', (2) 'either by the controller or *by any other person*.'[143] Here the judgment follows the objective, or absolute, approach to defining identifiability as the WP29 advances in the WP136.

The Court further moved to assess 'whether the possibility to combine a dynamic IP address with the additional data held by the Internet service provider constitutes a *means likely reasonably to be used* to identify the data subject'.[144] The Court took over the AG's argument that the possibility of combining a dynamic IP address with additional data *would not be reasonably likely* if it was 'prohibited by law or practically impossible' due to 'a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant'.[145] Although German law does not allow the Internet service providers to transfer data to the website providers, in the event of cyber attacks the website providers can contact a competent authority who can obtain information from the Internet service providers to start criminal proceedings. The website provider was found to have the means reasonably likely to be used to identify the website visitors on the basis of a dynamic IP address with the help of third parties, namely, the Internet service provider and the competent authority.[146] Hence, the dynamic IP addresses were found to be personal data in this case.

While agreeing with the WP29 on the outcome (both dynamic and static IP addresses are personal data according to the WP136[147]), the Court adopts a more measured approach to what *means of identification* are reasonably likely to be used. Namely, in addition to the cost and effort of identification that the WP29 includes among the factors that need to be taken into account, the Court considers the factor of legality. By itself, it would not have any restrictive effect on the scope of personal data compared to the WP136. Indeed, a legal possibility for the controllers to obtain additional information that enables them to identify data subjects is merely one of 'all the factors at stake' that should be taken into consideration to assess the possibility of identification according to the WP29.[148] The restrictive potential of *Breyer* lies in the assessment that starts from what would be *unreasonable*. The Court follows the AG's opinion that identification would not be reasonably likely if prohibited by law. This is in direct contradiction to the WP136. There, the WP29 names the possibility of organisational dysfunction, meaning also data security breaches resulting from illegal acts, among the relevant factors to be assessed. Indeed, identification can also result from illegal

[143]Ibid, [42] (emphasis added).
[144]Ibid, [45] (emphasis added).
[145]Ibid, [46].
[146]Ibid, [46]–[49].
[147]WP 136 (n 15), 16.
[148]Ibid, 15–16.

acts, eg hacking. Reflecting on the Court's argument, a more nuanced reasoning would be that a legal prohibition to combine data for identification would make the means of identification '*less reasonably likely* to be used', rather than '*not reasonably likely*'.

In addition, the Court does not consider the nature of processing in assessment of identifiability. Namely, under the WP136, the means of identification should be considered reasonably likely to be used when 'the processing … only makes sense if it allows identification of specific individuals and treatment of them in a certain way'.[149] It appears from the Court's reasoning in para. 47 that, under the circumstances of the case, registering IP addresses for the purposes of dealing with cyber attacks implies identification, as there is an established mechanism of identification of the perpetrators by their IP addresses for the purposes of bringing criminal charges where the Internet service providers, competent authorities, and the website providers collaborate. In brief, while the case presented an opportunity for the Court to confirm or reject the WP29 guideline 'implied identification = reasonably likely identification', the Court chose not to do so.

In sum, what is the impact of the *Breyer* case on the notion of personal data in part regarding identifiability? The case certainly reaffirmed the broad reading by the WP29 of 'all the means likely reasonably to be used either by the controller or *by any other person*'.[150] *Breyer* was the first case where the Court explicitly stated that it is not necessary 'that all the information enabling the identification … must be in the hands of one person'.[151] At the same time, the ruling that a legal ban on identification would make the identification means not reasonably likely to be used is a notable deviation from the WP136 and has a potential to limit a number of situations where data is considered identifiable in Europe. Crucially, this potential should not be overestimated. This is mainly because the outcome of the assessment of what the Court considered 'identification measures reasonably likely to be taken' was phrased by the Court in a manner specifically tailored to the narrow question posed, and the circumstances of the case. Citing the AG, it was not in dispute in this case 'whether the classification of dynamic IP addresses as personal data is necessary as soon as there is a third party, irrespective of who it may be, capable of using those dynamic IP addresses to identify network users'.[152] Answering the question posed this way negatively would indeed have significantly (and undesirably) limited the range of situations where information would be considered identifiable to a natural person, and hence personal data. However, the Court did not do so. Hence, the WP136 guidelines on identifiability have not been overruled and are still in play.

[149]Ibid, 16.
[150]*Breyer* (n 11), [42] (emphasis added).
[151]Ibid, [43].
[152]*Breyer* (n 11), Opinion of Advocate General Sanchez-Bordona General [50].

4.5. 'Any information': Nowak

Nowak, decided in December 2017, is significant for European data protection law in more than one respect. Among other things, it stipulates the conditions of the data protection rights of access and rectification.[153] The case is particularly important for the discussion of the concept 'personal data' since it provides interpretation for two out of three elements of the definition: 'any information' and 'relating to'. In this section I will focus on the former.

Until recently, the CJEU case law did not engage with the meaning of 'any information'. The Court dealt with this issue in *Nowak* for the first time. The core of the question before the Court was whether or not examination answers and examiner's comments constituted personal data within the meaning of the DPD, ie any information relating to an identified or identifiable data subject.

As to the meaning of 'any information', the Court ruled:

> the expression 'any information' […] reflects the aim of the EU legislature to assign a wide scope to [the concept of personal data], which is not restricted to information that is sensitive or private, but potentially encompasses all kinds of information, not only objective but also subjective, in the form of opinions and assessments.[154]

This interpretation is in line with the broad approach in the WP136 that any information can be personal data, regardless of its nature or content. While under the WP136 information can also be personal data regardless of format or medium (eg alphabetical, numerical, graphical, photographical or acoustic, kept on paper or in computer memory),[155] the Court is silent in this regard, and hence the WP136 remains the only compliance guideline on this matter. Nor did the Court address the gap in the WP136 and stipulate what information is. As a result, as the law stands, and depending on the theoretical perspective towards the meaning of information, everything can still be plausibly argued either to be or to contain information that can be personal data provided the other requirements of the definition are met.

4.6. 'Relating to': YS and others and Nowak

The Court so far has dealt with the meaning of information 'relating to' a person on two occasions. The first one was in 2014 in the joint cases *YS and others*.[156] There, in a situation where a broad interpretation of personal data would lead to what the Court saw as unintended results, it deviated from its earlier approach to interpret the scope of the Directive broadly and

[153] *Nowak* (n 24), [46] et seq.
[154] Ibid, [34].
[155] WP 136 (n 15), 7.
[156] *YS and others* (n 121).

consider proportionality of the resulting obligations as a secondary matter. As a result, the Court adopted a limited interpretation of 'relating to' compared to the WP136. *Nowak*, decided in 2017, effectively reversed this interpretation.

4.6.1. YS and others

At various stages of the immigration process in the Netherlands, YS, M and S, third-country nationals, requested access to the 'minute' in their immigration file on the ground of the right of access to their personal data. The 'minute' is a note drafted by an immigration case officer without authority to decide on an application for a residence permit to motivate his or her recommendation to a senior officer before he or she makes such a decision. The minute generally contains three types of information: (1) information relating to the case officer, ie name, telephone and office number; (2) 'data relating to the applicant', including name, date of birth, nationality, gender, ethnicity, religion and language; procedural history; statements of the documents submitted; and the applicable law; and (3) the 'legal analysis', ie 'an assessment of the foregoing information in the light of the applicable legal provisions'.[157]

Until 14 July 2009, it was the policy of the Dutch government to make a copy of the minute available upon request. Since then the policy has changed, and the applicant now receives 'a summary of the personal data contained in the document',[158] excluding the 'legal analysis'. YS, M and S were refused a copy of the full minute, and two sets of proceedings in Dutch courts have led to the references for a preliminary ruling by the Court of Justice. One of the central issues before the Court was whether the legal analysis in the minute should be considered personal data. If WP136 were followed, one would expect a positive answer: the legal analysis in the minute is information that is or is likely to be used '*with the purpose* to evaluate, treat in a certain way or influence the status or behaviour of an individual',[159] namely, the applicant for a residence permit. Yet, the Court answered negatively and concluded that 'although it *may contain* personal data, it *does not in itself* constitute such data'.[160]

The Court followed AG Sharpston and used two arguments to support its interpretation: first, 'a legal analysis is not information relating to the applicant … , but at most, in so far as it is not … a purely abstract interpretation of the law, is *information about the assessment* and application … of that law to the applicant's situation';[161] second, granting access rights in relation to a legal analysis is not compatible with 'the objective and general scheme'

[157]Ibid, [14].
[158]*YS and others* (n 121), [16]–[17].
[159]WP 136 (n 15), [10].
[160]*YS and others* (n 121), [39] (emphasis added).
[161]Ibid, [40].

of the DPD.[162] According to the Court, the data subject's rights in the DPD exist, among other things, so 'that person may be certain that the personal data concerning him are correct and that they are processed in a lawful manner', and the right of access exists to enable the data subject 'to carry out the necessary checks … to obtain, depending on the circumstances, the rectification, erasure or blocking of his data'.[163] Legal analysis 'is not in itself liable to be the subject of a check of its accuracy by that applicant and a rectification'.[164] Crucially, granting such a right of access 'would not in fact serve the directive's purpose of guaranteeing the protection of the … right to privacy with regard to the processing of data relating to him, but would serve the purpose of guaranteeing … a right of access to administrative documents, which is not however covered by Directive 95/46'.[165]

Consequently, the Court appears to interpret 'information relating to' narrowly, as information about an individual, and reject a broader approach of the WP136 where information can also relate to an individual not by virtue of its content, but by reason of the purpose or effect of its processing.[166] The Court does not refer to the WP136 in its judgment, or phrase the issue in similar words, ie as a choice between the information about an individual on the one hand and the information relating to an individual in purpose or effect on the other. However, it specifically rejects the argument brought on behalf of M and S that the legal analysis should be considered as relating to an individual 'if it is decisive for the assessment of the application … and is applied to the specific case of the applicant'.[167] Thereby the CJEU also rejects the WP136 understanding of 'relating to' in part of relation in purpose and effect. This is even more evident considering that the Court follows and explicitly refers to AG Sharpston's opinion. While not in the paragraph cited by the Court, AG Sharpston is clear that '[in her] opinion, only information relating to *facts about an individual* can be personal data. Except for the fact that it exists, a legal analysis is not such a fact'.[168] While the Court did not adopt this distinction between the facts and analysis, which would restrict the meaning of 'information relating to' even further, its narrow interpretation was likely to limit the range of situations within the scope of data protection law.

At the same time, *YS and others* did not shut the door completely for the broader interpretation of 'information relating to' in line with the WP136. In fact, the impact of *YS and others* can be restricted to similar situations only, ie where legal analysis is involved, and not affect situations which, while

[162]Ibid, [41].
[163]Ibid, [44].
[164]Ibid, [45].
[165]Ibid, [46].
[166]Part 3.4.
[167]*YS and others* (n 121), [35].
[168]Ibid, Opinion of Advocate General Sharpston [56] (emphasis added).

involving information used for assessment, are somewhat different from the facts in *YS and others*, as appears to be the case in *Nowak*.

4.6.2. Nowak

In December 2017 the Court revisited the meaning of 'information relating to' in *Nowak*. Similar to *YS and others*, in a case before the Irish Supreme Court Mr Nowak requested access to his examination script based on a data protection right of access to personal data. In its reference for a preliminary ruling the Irish court essentially asked whether the exam script containing the candidate's answers and the examiner's comments regarding those answers might constitute personal data.[169] Both the Advocate General in her opinion and the Court answered positively.

Since the Court followed the opinion of AG Kokott, reviewing the key arguments of the Advocate General is essential for understanding the Court's decision and its impact. The AG's argument links to the two-fold argument of the Court in *YS and others*. First, the AG examines what kind of a relationship between information and the individual is relevant in the context of the examination answers and the examiner's corrections to the answers. In both contexts the AG's reasoning is consistent with the WP29 approach to consider information as personal data when it is processed *with a purpose* to evaluate the status or behaviour of an individual. Although the answers are 'formulated in abstract terms or relate to hypothetical situations',[170] 'the script is a documentary record that that individual has taken part in a given examination and how he performed'.[171] While 'the extent of the link between an examination candidate and his performance in an examination' varies,[172] 'in every case, the aim of an examination […] is to identify and record the performance of a particular individual, ie the examination candidate. Every examination aims to determine the strictly personal and individual performance of an examination candidate'.[173] Therefore, contrary to the position of the Irish Data Protection Commissioner, the personal data in the script 'is not confined to the examination result, the mark achieved or even points scored for certain parts of an examination'.[174]

A similar assessment is made regarding the examiner's corrections on the examination script. The AG explicitly notes that *Nowak* resembles the case of *YS and others*,[175] and 'at the first glance' the Court's finding concerning legal analysis can be transposed to the examiner's corrections, namely that it is not information relating to the examination candidate, but 'at most information

[169] *Nowak* (n 24), Opinion of AG Kokott [2].
[170] Ibid, [19].
[171] Ibid, [21].
[172] Ibid, [23].
[173] Ibid, [24].
[174] Ibid, [27].
[175] Ibid, [58].

about the assessment'.[176] However, 'the purpose of comments is the evaluation of the examination performance and thus they relate indirectly to the examination candidate. … [C]omments on an examination script are typically inseparable from the script itself. Precisely because of that close link between the examination script and any corrections made on it, the latter also are personal data of the examination candidate […].'[177] Consequently, the AG adheres to the WP29's understanding of 'information relating to an individual' in terms of purpose twice in one case, with regard to the examination answers and the examiner's corrections.

As a second step in the argument, AG Kokott considers the relationship between the concept of personal data and the purpose of the right of access, the second pillar of the Court's position in *YS and others*. The AG does not expressly reject the reasoning in *YS and others* as she notes that the examination candidate 'has a legitimate interest, based on the protection of his private life' to be able to object to the processing of his or her examination script outside the examination procedure.[178] Yet her further statements are unequivocal: the purpose of the right of access under the DPD does not 'preclude the classification of an examination script as personal data';[179] 'the issue of right of access is only secondary' to the interpretation of the concept of 'personal data';[180] and 'the classification of information as personal data cannot be dependent on whether there are specific provisions about access to this information which might apply in addition […]. [N]either can problems […] with the right of rectification be decisive in determining whether there exists personal data'.[181] In other words, AG Kokott gracefully incorporated *YS and others* in her reasoning to the effect that the broad interpretation of 'the information relating to an individual' in terms of purpose stands.

The CJEU fully embraced the two-fold reasoning of the AG. First, the Court restated that the notion 'personal data' potentially encompasses any information, as long as it 'relates' to the data subject.[182] The latter condition is met where the information is linked to a particular person '*by reason of its content, purpose or effect*'.[183] This understanding is broader than the one offered by the AG and is a nearly *verbatim* adaptation by the Court of the broad interpretation of 'relate to' under the WP136.[184] In this case, the Court found the link between information and the examination candidate

[176]Ibid, [59].
[177]Ibid, [61]–[63].
[178]Ibid, [26].
[179]Ibid, [31].
[180]Ibid, [32].
[181]Ibid, [34].
[182]*Nowak* (n 24), [34].
[183]Ibid, [35] (emphasis added).
[184]WP29 adopts the test of 'content, purpose or result' (WP 136 (n 15), 11).

relevant as both the candidate's answers and the examiner's comments relate to the data subject in all three aspects: they reflect the information *about* the candidate (his knowledge, thought process and, in the case of a handwritten answer, information about his handwriting, as well as the examiner's opinion regarding the candidate's performance); the purpose of their processing is to *evaluate* the candidate in terms of his professional abilities; and the use of this information is 'liable to have an *effect* on his or her interests',[185] eg determine his or her employment chances.[186] When ruling that the examiner's comments constitute personal data of the examination candidate, the Court also did not consider it problematic that these comments relate to the examiner as well. As the CJEU explains, '[t]he same information may relate to a number of individuals and may constitute for each of them […] personal data', provided they are identifiable.[187]

As a second step in the reasoning, the Court adopted the argument of the AG as well and ruled that the status of information as personal data 'cannot be affected … by the fact that the consequence of that classification is, in principle, that the candidate has rights of access and rectification'[188] under the DPD. Indeed, when information is classified as personal data, the entire body of the data protection guarantees applies, which includes the data quality and security obligations and also access, rectification, and objection rights.[189] The examination candidate has a legitimate interest based on the protection of his private life, in exercising his or her rights of access, rectification, and objection under data protection law with regard to his or her answers and the examiner's comments. For instance, the candidate might object to processing of the answers and the examiner's comments outside the examination procedure,[190] or to ensure that the answers of another candidate were not ascribed to the candidate at hand, or that the answers of the candidate were lost, etc.[191] The data protection right of erasure might also be invoked and the answers and comments destroyed when the processing of the information in the identifiable form is no longer necessary for the purposes for which the information was collected, eg after the examination procedure is closed and cannot be challenged.[192] Therefore, giving a candidate access rights under data protection law serves the purpose of data protection law, ie 'of guaranteeing the protection of that candidate's right to privacy […], irrespective of whether that candidate does or does not also have such a right of access under the national legislation applicable to the examination

[185] *Nowak* (n 24), [39] (emphasis added).
[186] Ibid, [37]–[43].
[187] Ibid, [45].
[188] Ibid, [46].
[189] Ibid, [48].
[190] Ibid, [50].
[191] Ibid, [54].
[192] Ibid, [55].

procedure'.[193] The latter point is in direct contradiction to the Court's earlier judgment in *YS and others*. The Court explicitly notes this contradiction,[194] and thus de facto overrules *YS and others* in this part.

Read together with the opinion of AG Kokott, the *Nowak* judgement seriously limits the restrictive potential of *YS and others*, and allows for the further use of the wide interpretation of 'relate to' under the WP136.

To sum up, the Court generally supports the broad interpretation of identifiability in *Breyer*. The Court adopted a restrictive view on what 'information relating to' a person means in *YS and others*, suggesting that information only relates to an individual when it is about him or her. Yet in 2017 the CJEU effectively reversed *YS and others* and its limited reading of 'relating to' in *Nowak* and extended the interpretation to include relation by reason of content, purpose or effect. The Court also ruled that 'any information' is to be interpreted broadly to include information regardless of its nature or content. As to the format or medium of information or what 'information' is, the Court has yet to rule on these matters. If the Court follows its own case law, I do not expect that the judgments on these issues will be restrictive. Historically, the Court has considered the material scope of the DPD as very wide. It is not inclined to narrow the scope down because the legal rights and obligations that follow appear difficult to comply with, 'disproportionate' or 'unreasonable'. In fact, the Court regards those matters as secondary to the matter of scope. As a result, the broad interpretation of personal data under the WP136 still stands. To restate the devil's advocate's argument, weather can still be plausibly considered personal data in Europe.

5. Is there a problem?

Should European data protection law faithfully follow the WP29's broad interpretation of personal data, and should the Court in its future case law adopt a broad view on how information relates to people to warrant the application of data protection law? I will offer a nuanced answer here: from the perspective of the present and near future of data protection, I welcome the broad interpretation of the concept; yet, from the perspective of a more long-term future, while I still support the broad scope of the legal protection that a broad notion of personal data brings, I caution about the consequences of imposing the same high intensity of obligations in all data processing situations.

5.1. Present and short-term future under the GDPR

At present, to the question whether a broad concept of personal data is problematic and needs to be limited my answer is a definite no. This is based on

[193]Ibid, [56].
[194]Ibid, [56] ('a contrario, judgment of 17 July 2014, YS and Others, […] paragraphs 45 and 46').

two considerations. First, to restate the point made on a number of occasions in this article, if data has a potential to impact people, it should trigger some form of legal protection. The various elements of the concept 'personal data', especially 'relate to', as interpreted in the WP136 and the case law provide a roadmap towards mapping and assessing such intended and unintended impacts. If it appears that more data under more circumstances is likely to affect people, then more data should be considered personal and trigger application of data protection law. In this vein, personal data should be interpreted broadly.

Second, in the actual data protection practice today the range of situations that are considered processing of personal data are much more limited than the analysis in Part 3 might suggest, and we have not yet achieved the data protection 'system overload', as the next section will describe it. One reason for this is that the onlife world of total connectivity and data-driven agency – while firmly on its way – has not arrived and established itself yet. The technology of datifying everything and using any information to lay connections to people might be (nearly) there, but it has not yet been implemented in daily practice to the degree that is making everything personal data as we speak, albeit the share of the environment that is not converted into data and data that is not personal steadily shrinks.

In addition, while according to the law in the books the concept of personal data is broad and on the verge of becoming all-encompassing, in daily compliance practice the data that I would argue should be considered personal in the sense of the WP136 is often not recognised as such. This can happen because of a lack of awareness on the part of the data controllers about the exact meaning of personal data under data protection law. The first example that comes to mind is the discussion that data protection lawyers often have with data scientists regarding anonymisation. Data scientists assume that a certain data analytics project they plan has a clean data protection bill of health since the data in question is encrypted or stripped of identifiers and hence does not contain personal data. The lawyers respond that if the data scientists have or are reasonably likely to gain access to the encryption key or identifiers, the dataset in question is pseudonymous rather than anonymous, ie it does contain personal data and is subject to data protection law, leaving aside the fact that the encryption and pseudonymisation of the dataset also constitutes processing of personal data and is subject to data protection law. It is easy to imagine a similar discussion regarding the status of information as personal data when it is not 'about people' but relates to them in purpose or effect. The new GDPR is equipped with stronger enforcement tools than its predecessor, which hopefully will be instrumental in transforming data protection law in the books into data protection practice.

Another reason is that while the explanation of the concept of personal data in the WP136 is very comprehensive, it appears not comprehensive

enough as data processing technology and practice progress. As a result, there are a lot of future battles about the meaning of personal data that will have to be fought and won, in courts, before data protection authorities and in daily compliance practice, before even the data that already has impact on people is recognised as personal and triggers legal protection. Under these circumstances, narrowing down the concept of personal data is not the game one should play. I will name only a few examples of these future battles.

Given the ongoing merger of the material and online worlds, one of the transformations leading to the onlife world,[195] the Court of Justice may soon face the question of the meaning of information and how it relates to the information mediums, for instance, in cases involving human tissue samples. Even the meaning of identifiability, so far considered clear, might be broken open. While the WP136 provides elaborate guidelines on when a person is considered *reasonably likely* to be identified, the explanation of what 'identified' means is limited to 'a natural person can be considered as "identified" when, within a group of persons, he or she is "distinguished" from all other members of the group'.[196] Recital 26 names singling out as an example of what would constitute identification, leaving the rest to common sense. According to the GDPR definition of personal data, a person is identified '*in particular* by reference to an identifier'[197] [emphasis added]. This means that the definition also might include identification other than by reference to an identifier. At the same time, Leenes distinguishes four types of identifiability, where only one type is on the basis of an identifier.[198] The need to explore the meaning of identifiability further is likely to emerge in the context of facial detection technology used eg in 'smart' outdoor advertising boards which are now installed at railway stations, in shopping malls or near supermarkets. Such technology enables recognition of age, sex and facial expression of a person in front of the smart ad board, it can deduce the mood he or she is in at that moment and it can display in real time a suitable ad, tailored to a person or a group of people standing near that board. Because, in contrast to the facial recognition technology,[199] facial detection technology does not allow for the recognition of people again, its creators consider that the data processed is anonymous. The Irish Data Protection Commissioner in a recent press release took the view that

[195]Floridi (n 4) 2.
[196]WP 136 (n 15), 12.
[197]Article 4(1) GDPR.
[198]R Leenes, 'Do They Know Me? Deconstructing Identifiability' (2008) 4(1&2) *University of Ottawa Law & Technology Journal* 135.
[199]Facial recognition is 'automatic processing of digital images which contain the faces of individuals for the purpose of identification, authentication/verification or categorization of those individuals' (Article 29 Working Party Opinion 02/2012 on facial recognition in online and mobile services, 00727/12/EN, adopted on 22 March 2012 (WP 192), 2).

'the technology […] does not involve the recording, analysis, matching, profiling or storage of personally identifiable data'.[200] The question this case raises is whether displaying real time personalised ads to one or several people without the aim of recognising those people in future constitutes singling out and identification in the sense of the GDPR.[201]

Under these circumstances falling back on what has been agreed on as the meaning of personal data and narrowing it down would be counterproductive in light of the aim of data protection law to ensure effective and complete protection of data subjects.[202]

5.2. Long-term future of the GDPR

Also from the perspective of the long-term future of data protection, the all-encompassing interpretation of personal data and the resulting broad scope of legal protection by themselves should not be the problem. Indeed, one could argue that 'if the weather is going to be used to target and categorise me, I need protection against its potential to define me as dangerous or depressed', even if achieving this protection is difficult.[203] I agree. The problem as I see it is that in the circumstances where everything is personal data and everything triggers data protection, a highly intensive and non-scalable regime of rights and obligations created by the GDPR will not simply be difficult but impossible to maintain in a meaningful way. Two factors are key to the 'system overload': the broad personal scope of data protection law and the high intensity of data protection compliance under the GDPR, the latter weighing heavier than the former.

As to the personal scope, both the WP29 and the Court took a broad approach to the concept of the data controller, thereby widening considerably the range of actors subject to data protection obligations. The controller is the natural or legal person which alone or jointly with others determines the purposes and means of the processing of personal data,[204] and is the primary bearer of the data protection compliance burden. While a natural strategy of avoiding data protection obligations was to argue that an entity at hand does not have control over data processing and is a mere processor, the WP29 explained that the legally relevant control stems not only from the formal arrangements, such as law and contract, but also from the factual influence one has over data processing.[205]

[200]Data Protection Commissioner, *Press Release on the use of Facial Detection Technology in Advertising*, published on 15 May 2017, available online www.dataprotection.ie/docs/EN/15-05-2017-Statement-on-use-of-Facial-Detection-Technology-in-Advertising/i/1634.htm (accessed 17 February 2018).
[201]My answer would be yes, but I shall leave my argument for another time.
[202]*Google Spain* (n 116), [34].
[203]M Hildebrandt (in personal communication with the author).
[204]Article 2(d) DPD; Article 4(7) GDPR.
[205]Article 29 Working Party, Opinion 1/2010 on the concepts of 'controller' and 'processor', adopted on 16 February 2010 (WP 169), 11.

> This assessment allows for the drawing of external conclusions, assigning the role and responsibilities of controller to one or more parties. This might be particularly helpful in complicated environments, often making use of new information technologies, where relevant actors are often inclined to see themselves as 'facilitators' and not as responsible controllers.[206]

In another opinion, in the context of the social network providers ('SNS'), WP29 reasons that these are data controllers, since 'they provide the means for the processing of user data and provide all the "basic" services related to user management (eg registration and deletion of accounts)'.[207] An additional factor determining the status of the SNS as a controller is that SNS providers 'also determine the use that may be made of user data for advertising and marketing purposes – including advertising provided by third parties'.[208]

In *Google Spain*, the Court adopted a similarly broad approach to controllership in relation to the search engine providers and personal data uploaded on the third-party websites, arguing that a narrow interpretation would be contrary to the DPD's objective 'to ensure … effective and complete protection of data subjects',[209] even in a situation where personal data is uploaded on a third-party website and the controller has no knowledge of the data.[210] Notably, as decisive for establishing the factual influence over data processing, the Court cited the fact that 'processing [by the search engine] … can be distinguished from and is additional to that carried out by publishers of websites'.[211] This and the WP29's standards of the factual influence amounting to controllership could be applied to any entity with a business model consisting of providing somehow structured or new data processing services with a predetermined user interface, which makes it 'distinguished from and additional to' what their clients would do with the data otherwise. These entities would then be fully bound by the full scale of data protection obligations.[212] Yet the broad personal scope of data protection law is also not problematic by itself, which brings us to the next point, ie high intensity of compliance.

The current data protection regime under the DPD already establishes both *ex post* and *ex ante* controllers' obligations. For instance, prior to processing the controller has to define the purpose and a suitable legal ground of processing, which may include some impact assessment exercises, eg in assessing compatibility of processing with the original purpose of data collection, or if the controller considers using legitimate interest as a processing ground.

[206] WP 169, 11.
[207] Article 29 Working Party, Opinion 5/2009 on online social networking (WP 163), 5.
[208] Ibid, 5.
[209] *Google Spain* (n 116), [34].
[210] Ibid, [24], [34].
[211] Ibid, [35].
[212] Koops (n 35) criticises the GDPR for creating equal data protection regimes for different data processing situations.

The controller also has to assess legality, fairness and necessity of data processing before it commences, and monitor that the data is accurate and up-to-date throughout the lifetime of personal data. The GDPR relies on the culture of ongoing compliance even more heavily. To name just a few examples, think inter alia of the principle of accountability, which requires not only compliance with the GDPR but also the ability to demonstrate compliance;[213] data protection impact assessment implies ongoing assessment of the risks of data processing, and taking risk-mitigating measures before, during and after the data processing takes place.[214] These are mostly not just negative obligations to stay away from certain data processing (after all, there is no problem with general obligations to refrain from killing or stealing and these apply to everyone all the time); data protection law imposes positive obligations to constantly take active action to comply.

The broad material and personal scope of European data protection law combined with the high intensity of compliance under the GDPR, to exaggerate it ever so slightly, make data protection apply to nearly anyone processing nearly any information at nearly any time, and the threat of serious sanctions omnipresent. Going back to Lon Fuller, complying with and enforcing such a law in a meaningful way is impossible, and something – enforcement or compliance – will have to give. Facing the threat of effective and deterring sanctions,[215] the controllers, instead of engaging in a meaningful assessment of fairness and necessity, will be pushed to create the formal appearance of compliance by using 'compliance surrogates', such as compliance roadmaps, 'accountability tools,' trust marks and certification schemes. The latter will likely be provided by private entities – law firms, consultancies and audit firms – who are in the business of selling reassurances more than safeguarding data protection interests,[216] and whose products are bought to validate and broadcast compliance to the shareholders, data subjects and the data protection authorities. Data protection, using the words of Julie Cohen, will become 'Kabuki theater that distracts both users and regulators from what is really going on'.[217]

Alternatively, or in addition, enforcement of the data protection rights and obligations will likely become selective, determined by priority lists and shortcuts that the data protection authorities would develop to cope with the

[213]Article 5(2) GDPR.
[214]Art. 35 GDPR et seq.
[215]Think of the administrative fines in Art 83(4) GDPR of 'up to 10 000 000 EUR, or in the case of an undertaking, up to 2% of the total worldwide annual turnover of the preceding financial year, whichever is higher'.
[216]Eg N Purtova, 'Who Decides on the Future of Data Protection? Role of Law Firms in Shaping European Data Protection Regime' (2014) 28(2) *International Review of Law, Computers & Technology* 204.
[217]Julie E Cohen, 'The Biopolitical Public Domain: The Legal Construction of the Surveillance Economy' (forthcoming) *Philosophy & Technology*, available online at https://ssrn.com/abstract=2666570 (accessed 17 February 2018).

workload, including the 'compliance surrogates'. Selective enforcement and the pretence of compliance will reinforce each other, ridiculing data protection law, and depriving its protection of meaning.

6. Conclusion: what to do about personal data?

This contribution, having examined the notion of personal data in EU data protection law, concludes that in the onlife world of data-driven intelligence which is firmly on its way everything will soon fall within the definition of personal data as interpreted by the WP29 and the Court of Justice. The rapid progress of (re)identification technologies and failure of anonymisation certainly play an important role in this argument. However, the main contribution of this article was to argue that the other two elements of the definition, ie 'information' and 'relating to a natural person', are key to this shift as well. The legal definition of personal data is not based on any articulated understanding of information but must be interpreted broadly; this allows for the definition to embrace the theoretical perspectives which consider everything to be or contain information. Further, when increasing amounts of information are gathered from increasingly 'smart' environments that assess or affect us in real time, it is safer to assume that any information is likely to relate to a person.

Given that the onlife world of hyperconnectivity and data-driven agency has not yet arrived and firmly established itself, at present a broad definition of personal data does not cause problems. Indeed, it is to be welcomed in light of the aim of data protection law to ensure effective and complete protection of data subjects. Yet once the onlife world becomes a reality, the good intentions will backfire. The system of legal protection based on such an all-encompassing notion and high intensity of positive compliance obligations is not going to be sustainable in the long run. So what other options do we have for that future of data protection law and what should we do with the concept of personal data? Once the tipping point is reached and everything is personal data in the sense of the GDPR and the WP136, I see three possible ways forward.

One option would be to adopt a narrower meaning of personal data. However, this option has its own limitations. While the downside of having a too broad definition is stretching the scope of data protection law too far, the risk of opting for a (too) narrow definition is incomplete protection of rights and interests that data protection law was intended to provide. As the Court acknowledges, a broad concept of personal data and a very wide material scope serve the protective objectives of data protection law. Granted, what those objectives are is subject to debate,[218] and the wording

[218] G Gonzalez Fuster and S Gutwirth, 'Opening up Personal Data Protection: A Conceptual Controversy' (2013) 29 *Computer Law & Security Review*, 531; N Purtova, 'Default Entitlements in Personal Data in the Proposed Regulation: Informational Self-Determination off the Table … and Back on Again?' (2014) 30(1) *Computer Law & Security Review* 6.

of the GDPR does not provide much clarification, referring to 'the protection of the fundamental rights and freedoms of natural persons with regard to the processing of personal data'.[219] At the same time, cutting back on the range of situations to which data protection law should apply cannot be done simply because a broader interpretation is unpractical. It requires careful thought, both as to what the protective objectives are, and as to where the boundaries of 'personal data' should lie to achieve those objectives. Narrowing down the interpretation of 'relate to' to, for example, a relationship by reason of content would make little sense as it is not clear why the protection should be triggered by processing of information that is *about* a person, and not when the information is processed with the purpose of affecting or assessing a person, while the latter describes a significant part of those data processing situations which raise concerns. Likewise, it is doubtful that a dataset should be subject to the data protection regime when its owner has sufficient tools to identify people in it, and why it should be exempted when the database owner and the tools of identification are just a few transactions away from each other. In the age of software as a service, one does not need to be a 'digital giant' with one's own identification capacities, when one can easily 'borrow' those on a pay-per-use basis. Think of the powerful data analytics tools such as IBM's Watson[220] becoming increasingly available to a growing range of users.

Another option would be to preserve the broad interpretation of personal data but reduce the intensity of compliance obligations, for instance by matching the intensity to risks. Koops considered an option of having different sets of obligations for different kinds of personal data an improvement on the compliance regime under the GDPR.[221] Schwartz and Solove similarly proposed treating information with varying degrees of identifiability differently.[222] The GDPR arguably implements this idea, inter alia, in the qualified obligation to perform data protection impact assessment under Article 35, and the Article 11 exemption from the data protection obligations which require identification when such identification is not necessary for the purpose of processing. However, this option raises the same issue as the first one: how do we know where to draw the boundaries between the compliance regimes of different intensity, and how do we know a particular configuration will meet the protective objectives of the data protection law?

I believe that there is a third option which too deserves careful consideration, namely, to abandon the concept of personal data as a cornerstone of

[219]Eg Recital 10 GDPR.
[220]www.ibm.com/watson/.
[221]Koops (n 35).
[222]Schwartz and Solove (n 8) 1877.

data protection altogether, and seek remedies for 'information-induced harms' – understood broadly as any individual or public negative consequences of information processing – without a sentimental attachment to this familiar proxy. To preserve this formal distinction will always imply that there is also data that is not personal, and while the former triggers legal protection, the latter should not. This duality is at odds with the world where any information has the potential to affect people. Therefore, all automated information processing should trigger at least an obligation to assess what impact it is likely to have. In this sense, to abandon the formal use of the notion 'personal data' amounts to accepting that all data is personal. Some might argue that it is already a matter of good practice to perform such impact assessment prior to data processing. In fact, as I argued earlier in this paper, to meaningfully apply the definition of personal data and establish whether or not a planned instance of data processing is within the GDPR's material scope, one effectively has to assess both likely and factual intended and incidental impacts of data processing on people. However, the fact of the matter remains that there is no such legal obligation. The only way to change this is to shift the default setting to 'all data is personal' and use this departing point to build a system of scalable data protection rules.

Whichever option it is, it is essential for the success of future legal protection against 'information-induced harms' that it is grounded in a better understanding of information and its relationship to people, in well-understood and articulated problems that it intends to solve and in the fully mapped mechanics of the information-induced harms it intends to avert.

Acknowledgements

This contribution reports on the results of the project 'Understanding information for legal protection of people against information-induced harms' ('INFO-LEG'). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 716971). The paper reflects only the author's view and the ERC is not responsible for any use that may be made of the information it contains. The funding source had no involvement in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication. I am grateful to Prof. Ronald Leenes, Prof. Mireille Hildebrandt, Prof. Bert-Jaap Koops, Dr Frederik Borgesius and Mr Peter Hustinx for critically interrogating the argument developed in this paper either in personal discussions or in paper comments. All mistakes are mine.

Disclosure statement

No potential conflict of interest was reported by the author.

Funding

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 716971).

Notes on contributor

Dr Nadezhda Purtova (LLM CEU, MSc Leiden) is an Associate Professor at Tilburg Institute for Law, Technology and Society, the Netherlands.